

Grundkurs i statistisk teori, del 2

Räkneövning 5 - χ^2 -test och linjära modeller, 23.04.2015

1. Vid ett genetiskt försök studerade man den avkomma som erhöles vid korsning mellan två typer av marsvin. Av 87 ungar var 43 röda, 10 svarta och 34 vita. Enligt den genetiska modellen skulle dessa färger förekomma i proportionerna $9/16$, $3/16$ respektive $4/16$. Testa på signifikansnivå $\alpha = 0.05$ om den teoretiska fördelningen avviker från det experimentella resultatet.

$$\text{Test: } H_0 : (p_1, p_2, p_3) = (9/16, 3/16, 4/16)$$

$$H_1 : (p_1, p_2, p_3) \neq (9/16, 3/16, 4/16)$$

$$\text{Testvariabel: } Q = \sum_{k=1}^3 \frac{(O_k - E_k)^2}{E_k} \underset{\text{approx.}}{\sim} \chi^2(3 - 0 - 1)$$

$$\text{Testvärde: } q = 10.06$$

$$\text{Kritiskt värde: } \chi_{0.05}^2(2) = 5.99$$

$$q > 5.99 \Rightarrow H_0 \text{ förkastas på sig.nivå } \alpha = 0.05$$

2. Ledningen för ett företag vill undersöka hur arbetstagarna på tre olika avdelningar upplever sin arbetsmiljö. Resultatet från undersökningen kan sammanfattas i följande tabell:

Arbetsmiljö	Avdelning		
	A	B	C
Utmärkt	20	10	5
Tillfredsställande	30	35	10
Dålig	10	15	15

Testa på signifikansnivå $\alpha = 0.01$ om arbetsmiljön är oberoende av avdelning.

$$\text{Test: } H_0 : \text{Arbetsmiljö och avdelning oberoende}$$

$$H_1 : \text{Arbetsmiljö och avdelning ej oberoende}$$

$$\text{Testvariabel: } Q = \sum_{k=1}^3 \sum_{l=1}^3 \frac{(O_{k,l} - E_{k,l})^2}{E_{k,l}} \underset{\text{approx.}}{\sim} \chi^2((3-1)(3-1))$$

$$\text{Testvärde: } q = 15.22$$

$$\text{Kritiskt värde: } \chi_{0.01}^2(4) = 13.28$$

$$q > 13.28 \Rightarrow H_0 \text{ förkastas på sig.nivå } \alpha = 0.01$$

3. Vid en trafikundersökning på en av Finlands riksvägar räknades under 81 vardagar antalet bilar som passerade en viss plats under en viss tidsperiod. Följande resultat erhöles:

Antal bilar	0	1	2	3	4	5	6
Antal dagar	14	12	25	16	10	3	1

Testa på signifikansnivå $\alpha = 0.05$ om antalet bilar per tidsperiod kan anses vara Poissonfördelat. (Om $X \sim \text{Poisson}(\lambda)$ så är sannolikhetsfunktionen $p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$)

$X = \text{"antal bilar"}$

Test: $H_0 : X \sim \text{Poisson}$

$H_1 : X \not\sim \text{Poisson}$

λ okänd: Vi skattar med $\hat{\lambda}_{ML} = \bar{x} = 2.11$

$$\text{Testvariabel: } Q = \sum_{k=1}^6 \frac{(O_k - E_k)^2}{E_k} \underset{\text{approx.}}{\sim} \chi^2(6 - 1 - 1)$$

Testvärde: $q = 6.61$

Kritiskt värde: $\chi_{0.05}^2(4) = 9.49$

$q < 13.28 \Rightarrow H_0$ förkastas ej på sig.nivå $\alpha = 0.05$

4. För ett stickprov bestående av 10 parvisa observationer på variablerna (X, Y) har man beräknat:

$$\sum_{i=1}^{10} x_i = 12.00 \quad \sum_{i=1}^{10} x_i^2 = 18.40 \quad \sum_{i=1}^{10} y_i = 15.00 \quad \sum_{i=1}^{10} y_i^2 = 27.86 \quad \sum_{i=1}^{10} x_i y_i = 20.40$$

Anpassa regressionslinjen $y = \beta_0 + \beta_1 x$ till observationerna genom att skatta regressionskoefficienterna med MK-estimaterna $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ och $\hat{\beta}_1 = r_{xy} \frac{s_y}{s_x}$.

Först beräknar vi $\hat{\beta}_1 = r_{XY} \frac{s_Y}{s_X} = \frac{s_{XY}}{s_X \cdot s_Y} \frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2}$:

$$\begin{aligned} s_{XY} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \dots \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n} \right) \\ &= \frac{1}{10-1} \left(20.40 - \frac{12.00 \cdot 15.00}{10} \right) = 0.266\dots \end{aligned}$$

$$\begin{aligned} s_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \dots \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right) \\ &= \frac{1}{10-1} \left(18.40 - \frac{12.00^2}{10} \right) = 0.444\dots \end{aligned}$$

Vi får regressionskoefficienterna $\hat{\beta}_1 = 0.60$ och $\hat{\beta}_0 = 0.78$, dvs regressionslinjen $y = 0.78 + 0.60x$.

5. En grupp om totalt 60 studerande deltog i en undersökning där man jämförde tre olika undervisningsmetoder. Gruppen delades slumpmässigt in i tre lika stora undervisningsgrupper som därefter undervisades med de olika metoderna. Slutligen ordnades ett test för att jämföra resultaten för metoderna. Stickprovsmedelvärdena och -varianserna av resultaten i de olika grupperna beräknades:

$$\bar{x}_1 = 10.50 \quad s_1^2 = 8.83 \quad \bar{x}_2 = 9.73 \quad s_2^2 = 8.76 \quad \bar{x}_3 = 9.14 \quad s_3^2 = 11.17$$

Undersök om det finns någon signifikant skillnad i testresultaten för de olika metoderna. Välj signifikansnivå $\alpha = 0.05$ som ger det kritiska värdet $f_{\alpha}(k-1, n-k) = f_{0.05}(2, 57) = 3.16$.

Test: H_0 : Ingen skillnad i resultat mellan grupperna

H_1 : Skillnad i resultat mellan grupperna

Testvariabel: $F = \frac{SS_B/(k-1)}{SS_W/(n-k)} \sim F(k-1, n-k)$

Testvärde: $f = 0.97$

Kritiskt värde: $f_{0.05}(2, 57) = 3.16$

$f < 3.16 \Rightarrow H_0$ förkastas ej på sig.nivå $\alpha = 0.05$